Proceedings of the 17th International Conference on Automation & Computing, University of Huddersfield, Huddersfield, UK, 10th September 2011

Data Mining for Gearbox Condition Monitoring

M. Baqqar, M.Ahmed , F. Gu
School of Computing and Engineering
The University of Huddersfield
Huddersfield, HD1 3DH, UK
Corresponding author: M.Baqqar@hud.ac.uk

Abstract— **Engineering datasets have growing rapidly in size and diversity as data acquisition technology has developed in recent years. However, the full use of the datasets for maximizing machine operation and design has not been investigated systematically because of the complexity of the datasets and huge amounts of data. This also means that data analysis based on traditional statistic based methods are no longer efficient in obtaining useful knowledge from these datasets.**

Thus this paper discusses dynamic and static datasets collected from a gearbox test rig with a typical drive system such that the datasets are considered representative for condition monitoring purposes. Dynamic datasets were analyzed to diagnose the condition of the gear: Healthy or Fault, using conventional signal processing techniques such as time-domain and frequency-domain analysis. The static data was also analyzed for comparative evaluation of detection performances.

This procedure of data collection and analysis allowed a full understanding to be gained of condition monitoring datasets and paved the way for developing a more effective Data mining approach and efficient database.

Moreover, to evaluate the effectiveness of using these new techniques, a prototype database was developed based on a gearbox test system and tested using these methods. The results obtained from a number of conventional methods have shown that data mining can obtain information for condition monitoring efficiently but not so accurately to give fault severity information, which is often sufficient for making maintenance decisions.

Keywords- Gearbox condition monitoring, Data mining methods, Conventional methods .

I. INTRODUCTION

Condition monitoring (CM) is a technique for acquiring operating data and analyzing it to assess the health and condition of equipment. Thus potential problems can be detected and diagnosed at an early stage in their development, providing the opportunity to take suitable recovery measures before they become so severe as to cause machine breakdown. To obtain accurate results CM collects large amounts of data with wide diversity including operating parameters, high density dynamic signals and special event datasets to produce historical trends which are presented to engineers and stored in databases. This gives rise to the problem that the volume of data is very large and the relationship between measurements is very complicated. Consequently, the CM data is not always understood properly [1] and the extraction of useful and meaningful information from the data is extremely challenging. Because machines and sensor equipment are growing in complexity, combined with the recent progress in information technology (IT), data acquisition systems (DQS) can produce an overwhelming amount of data which is continuously increasing and contains features representing hundreds of attributes. Data mining (DM) techniques based on Computational Intelligence, Machine Learning (ML) and advanced statistics have created new techniques and tools for automated extraction of implicit, previously unknown and potentially useful patterns and knowledge.

DM is an Artificial Intelligence (AI) powered tool that can discover useful information within a database that can then be used as a guide to action. The use of DM techniques in industry began in the 1990s [2-4] and has steadily attracted more attention so that now DM is used in many different areas in manufacturing to extract useful information for use in predictive maintenance, fault detection, quality assurance, design, production, scheduling, and decision support systems.

II. GEARBOX CONDITION MONITORING

CM of a gearbox is a very important activity because of the importance of gears in power transmission in all manufacturing. Thus there has always been a constant pressure to improve the measuring techniques and analytical tools for early detection of faults in gearboxes. Gears are the most important element in the gearbox and, due to the high demands on them even at normal operating speeds and applied loads, gears are often subject to premature failure due to wear and material fatigue. To monitor the condition of a gearbox, physical parameters such as vibration, sound, temperature, even the motor current can be used.

Traditional methods of monitoring gearboxes are based on the assumption that any change in the condition of the gearbox may be detected by changes in the measured vibration signal [5]. This is because defects on a gear will alter both the amplitude and phase modulations of the gear's vibrations. Thus, any changes in vibration signal that can be measured and analyzed to provide an indication of the gearbox's condition.

The measured vibration signals often exhibit highly non-stationary properties, because the gear defects and incipient failures often show themselves in the form of changes in the spectrum of the signal. Thus, selection of the appropriate signal processing technique for detecting gearbox deterioration when the box is subjected to varying loads turns out to be crucial, since these vibration signals can also be heavily corrupted with noise, and are often non-stationary. The statistical parameters of the signal produced by the damaged component may not be altered much by the presence of a transient defect especially in the presence of noise while varying loads may give a signal in the form of unexpected or uncertain sources. Thus false alarms will be generated and unnecessary maintenance cost can be incurred [6].

III. CONDITION MONITORING USING DATASETS FROM A GEARBOX RIG

A. Data Characteristics

Gearbox data can be divided into two types: static datasets and dynamic datasets. As shown in the table, the static dataset contains mainly measurements from the controller and these were used to demonstrate the performance characteristics of the system. This type of data can give a quick indication of system health. The dynamic datasets are often used for CM as they can produce more details of the health condition and hence allow diagnosis of faults.

TABLE I. GEARBOX DATASETS

Static datasets	**Dynamic datasets**
Armature Current	Shaft speed
Load Set	Angular Speed
Speed Feedback	Motor current
Torque Feedback	Vibration signal from gearbox
Motor Current	Vibration signal from motor flange
Speed Demand	

B. Condition Monitoring Data Size

The amount of data acquired and stored for CM will be determined during the process of data collection via the data acquisition system. The memory size of the data collectors and data processors can affect the size of the collected data file. Data size and data points are very important factors in providing a clear picture of data collection. Table II shows size of the datasets collected from the gearbox test rig.

TABLE II. THE SIZE OF THE DATASETS COLLECTED FROM THE GEARBOX TEST RIG

Type of data	**Size of data file**
Static data	*364 KB (373,079 bytes)*
Dynamic data	*152 MB (160,016,680 bytes)*

C. Conventional Methods for Monitoring Gearbox

To better understand gearbox CM using traditional diagnostic techniques, experimental work was conducted on the same gearbox test rig as had originally been designed and previously used to monitor the health of a two-stage helical gearbox using conventional techniques of vibration monitoring. The two-stage helical gearbox was used for the experimental work because, in addition to its widespread use in industry, such gearboxes allow faults to be easily simulated.

The goal of the experiment was to monitor the gear under healthy conditions and compare the results with those obtained when a realistic fault was introduced under similar operating conditions. Generally, many types of fault can be observed in gearbox operation, a broken tooth, a crack, scuffing, wear, etc. In this experiment, one broken tooth was seeded onto the 1st gear because it is a very common fault in gearboxes, and data was collected under different loads.

Conventional techniques used in this work include the waveform comparison in the time domain and spectrum analysis in the frequency domain. In addition, a direct comparison is made of the static data based CM.

IV. STATIC MEASUREMENT BASED DETECTION

A. Detection from Static Datasets

From Figure 1, it can be seen that the load set was similar for both healthy and faulty conditions. There was a slight difference in the armature current but this difference is not sufficiently significant to be considered as a fault indicator.

Torque feedback gives the best indication of the presence of the fault, and it can be seen from the figure that there is an increase in the Torque feedback signals.

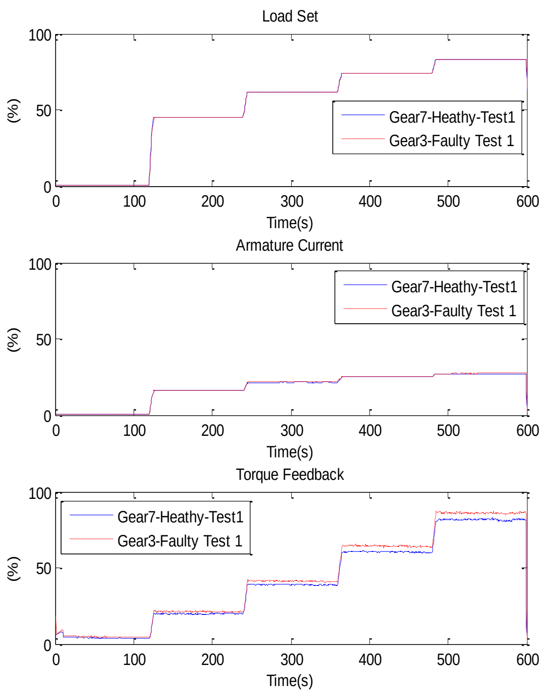

Figure 1. Load set, Armature current and Torque feedback for healthy and faulty gear conditions

From Figure 2 and 3, comparing the two conditions it can be seen that there is a slight difference between the healthy and faulty condition for all measurements.

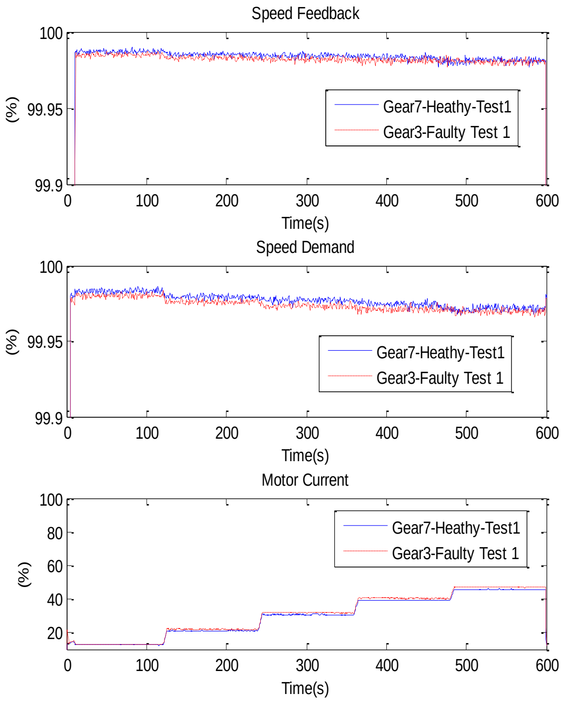

Figure 2. Speed Feedback, Speed Demand and Motor Current conditions for healthy and faulty gear

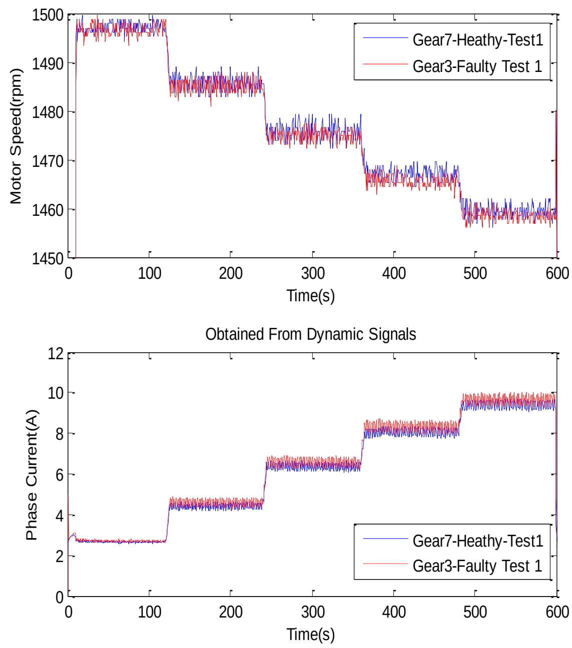

Figure 3. Motor Speed and Phase Current obtained from Dynamic Signal

V. VIBRATION BASED DETECTION

A. *Detection in the Time-domain*

Vibration waveforms for a healthy and faulty gear were collected from the accelerometer mounted on the gearbox casing with different current load operating conditions as shown in Figure 4. It can be seen that there are certain differences between the signal amplitudes due to the load variations for both gear conditions. Additionally, the waveform of the signals contains a massive amount of unknown information.

By comparing the two conditions it can be seen that all the waveforms of the signals exhibit some distortion but there are no clear fault indications even under high load conditions.

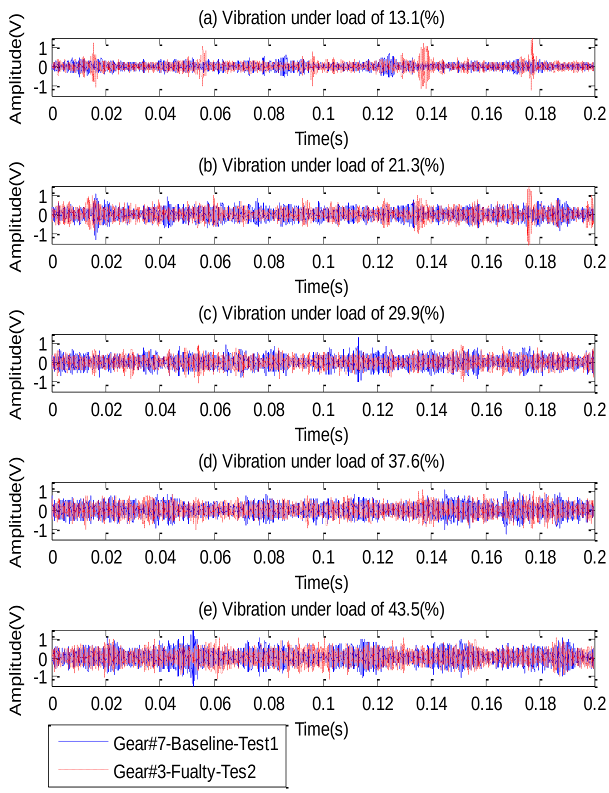

Figure 4. Vibration signals for healthy and faulty gears

For a more accurate clear and reliable assessment of the impact of load variation on the faulty gear, the RMS, Kurtosis and Peak values were found for the vibration signals.

From Figure 5, it can be seen that these statistical parameters varied with load, but there is no immediate or clear pattern to the fluctuations.

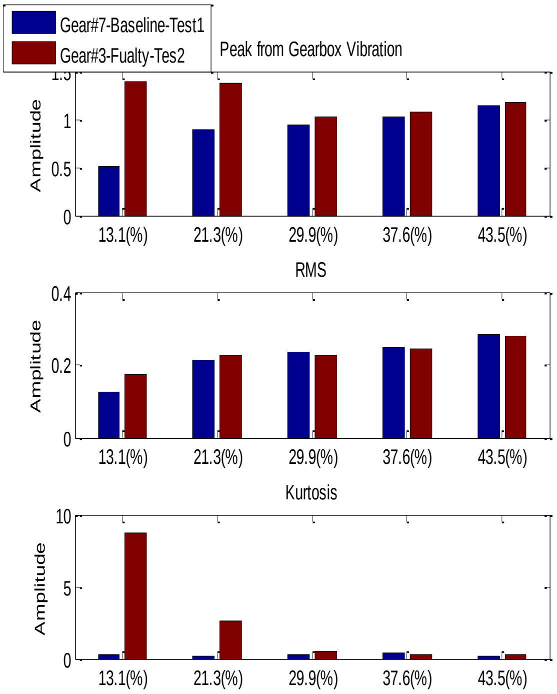

Figure 5. Peak, RMS and Kurtosis of time-domain vibration signal from gearbox

B. *Detection in the Frequency domain*

Frequency domain analysis for vibration signals of the healthy and faulty gear was carried out using the Fast Fourier Transform (FFT). The full spectrum of the vibration based on averaged healthy and faulty vibration signals are shown in Figure 6. Extracted from the figure are the dominant components in the spectrum are the transducer resonance responses.

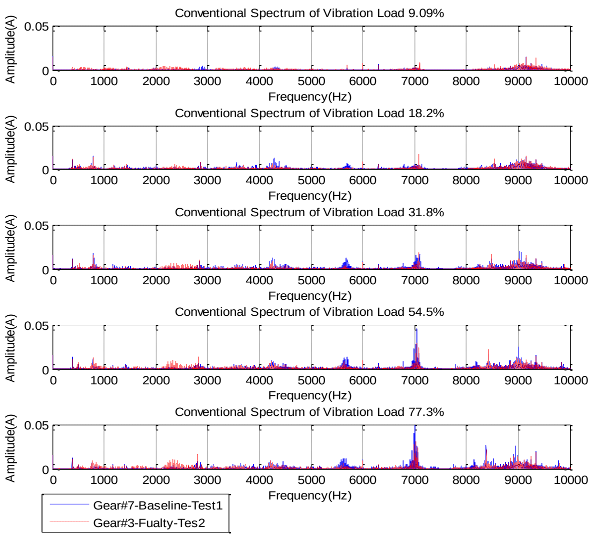

Figure 6. FFT spectral analysis of vibration signals from healthy and faulty gearbox

VI. CURRENT BASED DETECTION

A. *Detection in the Time-domain*

Figure 7 shows distortion of the motor current signal – difference from a pure sine wave - but no clear symptoms of a fault appeared even under different loads.

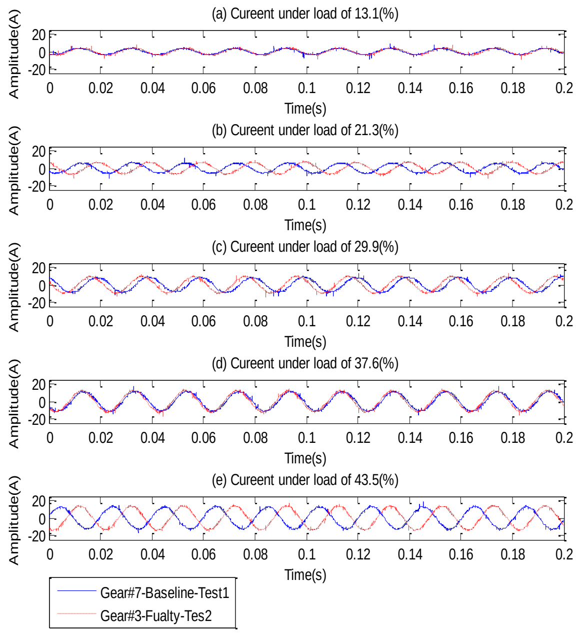

Figure 7. Stator current signals for healthy and faulty gears

From Figure 8, it can be seen that for all loads the Peak, RMS and Kurtosis for motor current are not significantly different for both healthy and faulty conditions.

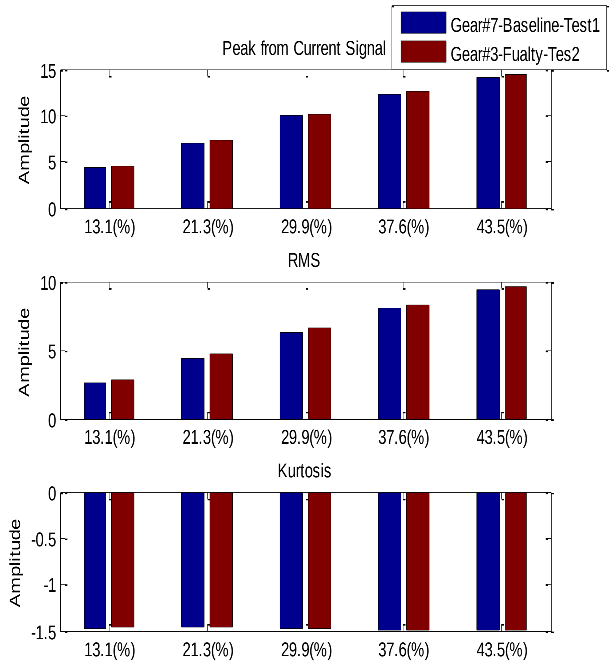

Figure 8. Peak, RMS and Kurtosis of stator current signals

B. Detection in the Frequency- domain

It can be seen from Figure 9 that the vibration spectrum is rich with discrete frequency components (0 to 100Hz).

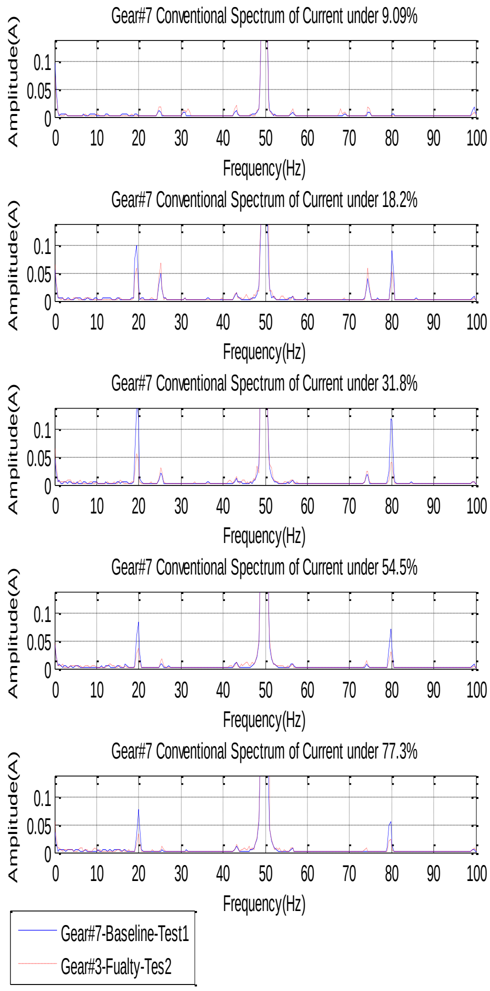

Figure 9. Spectral analysis of stator current signal measured for healthy and faulty conditions

VII. CONCLUSIONS

The growing volume of CM data challenges conventional CM technologies and techniques in obtaining cost effective results. In parallel, DM technologies include neural networks, evolutionary algorithms, pattern recognition and support vector machines are also developing rapidly due to the advances in computing hardware. It is important to apply these technologies to the accumulated data to achieve more accurate and efficient CM.

The analysis of the dynamic datasets using conventional methods such as feature extraction from the time- and frequency-domains shows good detection and diagnosis results. However, the amplitude of the features is not sufficiently high for reliable diagnosis.

The primary study of the static data also shows certain detection information but not enough to identify the root cause of the faults.

VIII. FUTURE WORK

• Pre-processing the dataset (choosing of data resources, cleaning the data from noise and errors, treatment unknown values, projection and reduction of data, etc.).

• Evaluating intelligent data mining techniques such as GA, NN, and SVM for Gearbox data.

• Selecting the DM task and technique for the extraction useful information and interesting and frequently occurring patterns.

• Build a mathematical model for vibration signal characterization under healthy and faulty gear condition; this requires the application of a suitable (mathematical and intelligent) data mining algorithms (NN), which describes the patterns.

• Validate the results of modelling based on the experimental data.

• Evaluation (test model, interpret and evaluate model).

• Post-processing the discovered or extracted patterns (further selection, ordering or elimination of patterns, visualization of the results).

• (Frequently some of the data mining DM steps need to be iterated several times to finally attain this goal).

REFERENCES

[1] S. McArthur; S. Strachan; and G. Jahn; "The design of a multi_agent system for transformer condition monitoring," IEEE Transactions on Power System, vol. 19, no. 4, pp. 1845_1852, 2004.

[2] Lee, M. H , 1993, "Knowledge Based Factory," Artif. Intell. Eng, 8, pp. 109_125.

[3] Irani, K. B , Cheng, J, Fayyad, U. M., and Qian, Z , 1993, "Applying Machine Learning to Semiconductor Manufacturing," IEEE Expert, 8(1), pp. 41–47.

[4] Piatetsky-Shapiro, G., 1999, "The Data Mining Industry Coming of Age" IEEE Intell. Syst, 14(6), pp. 32_34.

[5] Stander, C. J., Heyns, P. S. and Schoombie, W. 2002. Using Vibration Monitoring for Local Fault Detection on Gears Operating Under Fluctuating Load Conditions, Mechanical Systems and Signal Processing. Vol. 16, No. 6, pp. 1005-1024.

[6] Zhan, Y., Makis, V. and Jardine, A.K.S., 2006. Adaptive State Detection Of Gearboxes Under Varying Load Conditions Based On Parametric Modelling, Mechanical Systems And Signal Processing, Vol. 20, No. 1, pp. 188-221.